News Analysis and Summarization from RSS feeds

Supratik Chatterjee
B.Tech, CSE, Third Year,
SRM IST,
Vadapalani, Chennai,
Tamil Nadu, India.
RA1511003040196

Shubham Bagadiya
B.Tech, CSE, Third Year,
SRM IST,
Vadapalani, Chennai,
Tamil Nadu, India.
RA1511003040258

Sanjeev Siva
B.Tech, CSE, Third Year,
SRM IST,
Vadapalani, Chennai,
Tamil Nadu, India.
RA1511003040193

Abstract

In the field of media, news media is nowadays affected greatly by the social media. It becomes necessary to understand the facts of an event and to quantify or relate them to understand it. Most humans do not have the dedication to study all events at all the time. However, a computer could do it for them.

The aim of the project is to centralize information collected from news websites for connected and summarized reading. This aims to remove co-occurance of facts, and find discrepancies of figures, so that one might deal with them as an outlier. The rest of the data must be catalouged and indexed such that they may be referenced easily. When this is hooked up with a Machine Learning tool, should be able to predict outcomes, provided the dataset given to it has been collected for a duration long enought for it to work on developing itself.

Collection of information has to be from the several RSS feeds found online. Using those as the data point one can fetch the information and centrlize the resources for further analysis and prediction. This project will be an exercise in centrality, distributed computing, natural language processing, data mining, clustering and machine learning.

Index Terms

Natural Language Processing, Web Scraping, Text Mining, news, Co-occurence Analysis, Information Centralization, Machine Learning, Clustering, News Aggregation